DOCUMENT RESUME

ED 088 390 HE 005 279

AUTHOR Peterson, Richard E.; Vale, Carol A.
TITLE Strategies for Assessing Differential Institutional Effectiveness.
PUB DATE Oct 73
NOTE 25p.

EDRS PRICE MF-$0.75 HC-$1.85
DESCRIPTORS Educational Research; *Higher Education; *Program Effectiveness; *Student Development; *Student Experience; *Student School Relationship

ABSTRACT

This paper is an attempt to respond to the need for workable procedures for assessing the effectiveness of programs and institutions in complex postsecondary education systems. Effectiveness is taken to mean the capacity of the institution to advance student development-academic, vocational, and affective. Four general strategies are outlined in the paper that are intended to: (1) yield information directly applicable to policy issues and decisions; (2) yield information in a timely fashion; and (3) be implementable, in the sense of practical feasibility. The four plans briefly outlined, are the following: (1) Senior Assessment; intellectual competence focuses on the development of a number of intellectual and academic attributes while taking into account the general academic ability of the student at the time he enters as a freshman. (2) Sophomore Assessment: intellectual and/or vocational: affords a method for evaluating institutional effectiveness in either general education or vocational training programs during the first two postsecondary years. (3) Cross-sectional: intellectual and nonintellectual, generated by all the information yielded by 1 and 2 as well as intellectual criteria. (4) Alumni survey permits taking freshman ability into account in assessing postgraduate achievement. (Author/PG)

ED 088390

October 1973

STRATEGIES FOR ASSESSING DIFFERENTIAL INSTITUTIONAL EFFECTIVENESS IN MULTI-CAMPUS SYSTEMS

Richard E. Peterson and Carol A. Vale

Educational Testing Service, Berkeley

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

HE 005 279

Why be concerned about evaluating the effectiveness of colleges and universities? Isn't it commonly assumed that most higher education institutions are doing their job well enough--that students are being prepared satisfactorily for occupational careers, are acquiring an awareness of their cultural heritage, are developing as responsible citizens? Very likely, most colleges <u>are</u> doing a reasonably--perhaps minimally--satisfactory job, that is, are reasonably effective in some sense. However, there is growing awareness among both higher education professionals and informed representatives of the public that some kinds of higher education programs may be more effective than others, and especially that some may be more <u>cost-effective</u> than others. This general issue has taken on particular significance in the past decade with the emergence of dramatically new forms for teaching and learning--student-devised curricula, abolition of many traditional graduation requirements, various off-campus ("real world") learning experiences, cluster colleges, and so forth.

How might institutional "effectiveness" or institutional "quality" be defined? In a recent special report, the Carnegie Commission on Higher Education (1973) stressed the notion of "value-added:"

> The quality of an institution should be determined by what it does for the students it enrolls, not by the characteristics of its entering students...(p. 39).

Thus the fundamental index of institutional quality/effectiveness for most

colleges and universities would be how much the student learns or otherwise develops as a result of attending the institution. The Carnegie Commission goes on to underscore the importance of <u>differential</u> analyses:

> With this definition, the state college enrolling large numbers of freshmen from the middle of their high school graduating class has just as great an opportunity to achieve excellence through its vocational and academic programs as the more highly selective state university or liberal arts college...Does a residential college do a better job than a nonresidential college?...Only with a new definition of institutional quality and a means to measure it will parity of esteem become possible and universal access a reality (p. 39).

Effectiveness, as it is discussed here, is defined (only) with reference to goals; institutional effectiveness (quality) means, in short, achievement of institutional goals. We recognize that different colleges (e.g., in a state) may have markedly different goals. At the same time, colleges in a multi-campus system would typically have a core of goals in common. Furthermore, a single campus can be differentially effective for different types of students (e.g., students of differing intellectual ability).

Thus, the term "differential" will apply to assessment of:

(1) The effectiveness of different campuses (possibly to some extent pursuing different goals) within a single statewide system;

(2) The effectiveness of different systems (again pursuing common and different goals) within a total state post-secondary education complex;

(3) The effectiveness of a given institution (and its departments or programs) for different types of students enrolled.

None of the assessments to be outlined will be inexpensive to carry out. Yet the payoff in the form of identification of educational strengths and weaknesses in the system, assuming the project is conducted resourcefully and with integrity, should be worth the expense and effort. For in the final

analysis, the fundamental reason for undertaking the assessment in the first place is improvement of the effectiveness of the total system.

* * * *

While there has been much discussion in the past several years about the importance of assessing quality, measuring outcomes (or output, or productivity), and the like, the comment has tended to be hortatory and quite general. Our purpose in this paper is to deal with the topic more concretely, by presenting, in nontechnical language, four general analytic approaches to appraising differential institutional effectiveness in multi-campus systems.[1] Effectiveness will be taken broadly to mean the capacity of an institution (or entire system) to advance student development--academic, vocational, and affective.[2]

Each assessment strategy is intended to:

(1) Yield information directly applicable to policy issues, questions, and decisions;

(2) Yield information in a timely fashion;

(3) Be implementable, in the sense of the practical feasibility of carrying out the assessment project.

In brief summary, the four plans are the following:

Plan A. Senior Assessment: Intellectual Competence. This plan focuses on the development, after four years of college, of a number of intellectual and academic attributes, while taking account of the general academic ability of the student at the time he enters as a freshman.

Plan B. Sophomore Assessment: Intellectual and/or Vocational. Plan B affords a method for evaluating institutional effectiveness in either general (and transfer) education or vocational training programs (or both) during

the first two post-secondary years (again, accounting for differential ability level).

Plan C. Cross-sectional: Intellectual & Nonintellectual. Plan C generates all the information yielded by Plans A and B (which focus on intellectual criteria), plus a cross-sectional description of entering freshmen, end-of-year sophomores, and graduating seniors on nonintellectual (affective) as well as intellectual criteria.

Plan D. Alumni Survey. The criteria of effectiveness in this plan are various achievements and activities on the job, in graduate school, or elsewhere for alumni two years after graduation (from either a two- or four-year institution). As with the other plans, this one also permits taking freshman ability into account in assessing post-graduate achievements.

Depending on the kinds of issues and questions for which data are needed, one or more, or various combinations, of the four suggested plans could be followed. All possible information--the most comprehensive assessment envisioned in the four plans--can be achieved through a combination of Plans C and D. The minimum assessment for a system of four-year institutions would be Plan A; the minimum for community colleges, Plan B. An intermediate approach might be a combination of Plan A (or B, for two-year colleges), plus some or all aspects of the nonintellectual component of Plan C.

The plans as set forth are regarded as general outlines, to be modified and adapted according to system interests and resources. In particular, all assume flexibility in the choice of assessment (criterion) variables; that is, while general kinds of measures will be suggested, the specific concepts/instruments would be selected by the assessment project staff.

Choice of instruments, of course, is a critical element in the overall assessment. Published standardized tests have the advantage of being "known

quantities," of having national norms, and of being reliable and otherwise technically well-constructed. Locally (system) developed instruments have the advantage of being potentially more relevant to system and campus educational goals (these could be criterion- or performance-based, in the sense of mastery of specific curricular objectives). In any event, the several student questionnaires called for in Plans A, B, and C must be specially constructed. Likewise, there is no published standardized alumni questionnaire (Plan D) suitable for large-scale use.

It goes without saying that in managing any of these assessments, system staff will need to give thought to the most appropriate division of labor and resources between central office and component campuses. Thus, the very difficult process of deciding on criterion variables and specific measures could be accomplished jointly, cooperatively, as could construction of certain needed instruments, and preparation and review of the final report (data to be included, organization of tables, nature of interpretation, etc.). Selection of samples (following guidelines developed jointly), conducting the assessment(s), locating addresses of alumni, and assembling pre-freshman ability (control) scores could be done by the campuses. Data processing and analyses, on the other hand, might best be handled centrally.

It is important to emphasize that no <u>implementable</u> plan for assessing differential effectiveness can be conceived for which conclusions are not somewhat jeopardized by the nonrandom "assignment" of students to institutions. That is, different institutions attract and often select different types of students. Even if the bases for institutional and self-selectivity were specifiable, the degree of their relationship to various dimensions of effectivenesss is not straightforwardly determinable. This, together with the possibility that certain campuses may be more effective for some kinds

of students than for others, makes statistical "corrections" for student input differences inappropriate.[3] Our suggested approach, then, involves comparison of comparable types or categories of students across all institutions wherein the type is represented, on whatever dimensions of student output are deemed appropriate. This will require categorizing students on the basis of whatever input information is available or can be collected retrospectively, and we will recommend, at a minimum, pre-freshman general academic ability (tested) as the basic input control variable. This procedure does not deal entirely with the problem of nonrandomness, but to the extent that the student characteristics most relevant to the criteria of effectiveness are known and measurable, the inevitable compromise of experimental rigor with the demands of practicality becomes steadily more comfortable.

While the assessment strategies will seem to emphasize description of outcomes, all four plans have potential for identifying reasons why some institutions/programs are more effective than others. The fruitfulness of such "diagnostic" work would be limited only by the resourcefulness of the assessment staff in soliciting information bearing on the "fit" between student and institution/program characteristics. The significance of (additional) efforts to pinpoint causes of differential effectiveness is not to be minimized, when one considers, again, that the purpose of the assessment is institutional/system improvement.

In the pages that follow, we first present a summary table (Table A) which outlines the chief components of all four plans together with illustrative variables and measures. Each plan is then discussed separately, from the standpoints of (1) general purpose and logic, (2) illustrative policy questions answerable from the assessment, and (3) steps involved in conducting the assessment, together with a suggested time schedule.

TABLE A: SUGGESTED SPECIFICATIONS FOR FOUR HIGHER EDUCATION SYSTEM ASSESSMENT PLANS

Designation	PLAN A Senior Assessment: Intellectual Competence	PLAN B Sophomore Assessment: General and/or Vocational Education	PLAN C Cross-sectional: Intellectual & Nonintellectual	PLAN D Survey of Recent Alumni
Students Assessed	Graduating seniors	Sophomores (completing two full-time years at two- or four-year institutions)	Entering freshmen Sophomores (same as Plan B) Graduating seniors (same as Plan A)	Alumni (two years after receipt of degree or certificate)
Number	All, or sample of 2,000	Same as Plan A	Same as Plan A; same N all three groups	Same as Plan A
Criterion Variables (and illustrative measures)	General knowledge (UP Area Tests, SCA) Specialized knowledge (UP Mini Field Tests) Intellectual disposition (OPI scales) Satisfaction with college (CSQ Satisfaction scales) Other information (specially designed Senior Questionnaire)	General education (and community college transfer) students: Same as Plan A, except no assessment of specialized knowledge Vocational students: Self-report perceptions of quality of: instruction, equipment, program organization, campus climate for vocational educ., employment prospects (specially prepared Vocational Students Questionnaire) Basic skills: writing, mathematics (STEP II, SCA)	Intellectual Freshmen: same as Plan B-general/transfer Sophomores: same as Plan A Seniors: same as Plan A Nonintellectual (same at all three class levels) Autonomy, Personal Integration, etc. (OPI) Cultural Sophistication, Liberalism, Social Conscience (CSQ) Self-Actualizing Value, Self Regard, Time Ratio, etc. (POI) Locus of Control (Rotter I-E Scale) Current Affairs Knowledge	Self-reported: Employment situation Job satisfaction Earnings Graduate school situation Reasons for enrolling in the particular graduate school Continuing education Avocational activities Feelings about undergraduate experience Suggestions for improving undergraduate education Community activities Various attitudes
Testing Time	3 hours	2-1/2 hours (both general/transfer and vocational students)	5 hours; freshmen, 4-1/2 hours	Average time to complete survey questionnaire: 45 minutes
Input Ability Control Variable	SAT V+M, or ACT Composite, or score on other entrance test standard in the system	Same as Plan A, or equated scores from other tests (e.g., SCAT, CQT, CGP, CTAA)	Intellectual Same as A or B Nonintellectual None required	Same as A or B (depending on whether institution is two- or four-year)
Analytic (breakdown or "blocking") Variables	Freshman ability (four levels), only, or in combination with: sex major field socioeconomic background native/transfer (All from Senior Questionnaire)	Same as Plan A (except major field to include vocational program, and no native/transfer breakdown) (All from Sophomore Questionnaire and Vocational Student Questionnaire)	Freshman ability Sex Major field Socioeconomic background (All from a standard specially prepared Student Questionnaire, which may include some or all of the nonintellectual measures)	Same as A or B (depending on whether institution is two- or four-year)
Basic Statistical Methods	Analysis of variance of mean scores on criterion measures for seniors blocked according to freshman ability (four levels); and for sex by ability, major field by ability, etc. Possible use of multivariate procedures (e.g., discriminant, canonical, factor analyses) to examine patterns among criterion variables and institutional characteristics.	Same as Plan A Some or all data from Vocational Student Questionnaire analyzed via frequency (and percent) tabulations and chi-square tests.	Analysis of variance of criterion variable means for freshmen, sophomores, and seniors--for the total class-groups, and for the groups variously blocked. Frequency tabulations and chi-square for questionnaire items Possible multivariate procedures (per Plan A) Separate analyses for drop-outs	Frequency (and percent) tabulations for respondents blocked according to freshman ability; and for sex by ability, major field by ability, etc. Chi-square tests of differences among frequency distributions in various blocks (cells) Possible multivariate procedures (per Plan A)

Plan A. Assessment of Seniors: Focus on Academic/Intellectual Competence

This assessment plan consists essentially of comparisons of graduating seniors at a set of four-year institutions on designated academic and intellectual dimensions. Various tests of academic learning, as well as selected nonachievement measures (e.g., intellectual attitudes, styles, commitments, satisfaction with various elements of the college experience) are suggested as components of the assessment criteria. The contrasts among campuses would incorporate an index of general academic ability at the time of college entry, in a manner which permits conclusions about differential effectiveness for students of differing levels of ability. Other policy-relevant "breakdown" variables may also be used, such as academic field, socioeconomic background, off- vs on-campus residency, and the like.

Some of the kinds of questions that could be answered from the Plan A assessment include the following:

(1) Are there differences among campuses in the level of general knowledge of graduating seniors? In their intellectual commitments? Are there differences between multi-campus systems--public or private--on any of these indices of effectiveness?

(2) Are there differences in level of specialized knowledge for graduates in the corresponding disciplines across the campuses? For example, do biology graduates know more about biology at campus X than at campus Y?

(3) What are the patterns of satisfaction with various aspects of the college experience? By major field? By campus? By system?

(4) What is the pattern of differences across campuses (and systems) for students of a given ability level? That is, are some campuses (programs) particularly effective for students of modest ability? For students of high ability? (This is the general question of "value added.")

(5) What are the relationships between certain student background factors--e.g., socio-economic level, age, sex--and the various (intellectual competence) criteria of effectiveness? For the system? By institution? By major field?

(6) How do transfer students (of comparable ability) compare with native students in academic achievement at the time of graduation?

(7) What institutional/program characteristics are associated with high (or low) academic achievement? With intellectual commitment? With satisfaction with college?

A detailed study plan cannot be specified in advance of decisions delimiting its scope and objectives; the outline below, however, indicates the major steps involved, with a possible time schedule in the right margin.

(1) Determine the criterion variables and specific instruments for assessing each. Table A presents a suggested set of variables and instruments, which is to be regarded only as illustrative. Other (comparable) instruments used in an ongoing program of senior testing within a system could be appropriately substituted. — October through March

(2) Prepare Senior Questionnaire. This would include information to be used in the data analyses as breakdown variables (e.g., background factors, major field, etc.), as well as criteria not covered by the standard tests (e.g., original and present educational goals, future plans, etc.). It would require no more than 1/2 hour to complete. — October through March

(3) Determine information to be used for control of differential input. Certain kinds of data must be available for appropriate accounting of different levels of student ability or academic preparedness. Preferably, there would be standard systemwide pre-admission scores (on the SAT, ACT, or some comparable test). There is no requirement, for the kind of analysis of differential effectiveness proposed here, that the same test data be available at both freshman and senior levels. — October through March

(4) Design data management procedures. Computerize procedures for merge of pre-admission and senior data, with provisions for identification of dropouts and untested or incompletely tested seniors. Design and test management system and articulate with analytic (statistical) programs, the latter to be adapted or written, as required. — October through March

(5) Conduct the assessment. Unless the senior student population is extremely large, it would be necessary to test it in its entirety in order to have sufficient numbers of students in each of the proposed breakdowns. If sampling is possible, a stratified random sampling plan would be designed to ensure adequate coverage of all elements of the student population relevant either to matters of policy or to performance on criterion measures. — April

The amount of testing time would depend upon the criteria and measuring instruments used (not to exceed 5 hours), and would ideally be scheduled for a single session with a 1/2 hour break. Large group

testing situations would provide the most efficient coverage of the student population, and a required rather than a volunteer or persuasion procedure should be followed.

(6) Process data. Score standard instruments and transcribe questionnaire responses. Merge with pre-admission data tape; create master file. — May-June

(7) Analyze data. Various sorts of analyses are possible, and which would be done will depend upon the kinds of questions judged important by the system and assessment staff. — July through October

Comparisons of effectiveness among institutions and/or groups of institutions, with appropriate blocking on variables which are policy-relevant or related to the criterion, may be made by standard higher order analysis of variance (ANOVA) techniques. The blocking strategy will permit assessment of differential effectiveness for students classified along the blocking dimensions and allow detection of particularly good (or bad) matches of student type (e.g., ability level) with institution.[4]

Effectiveness criteria assessed in frequency form (on the Senior Questionnaire) would be analyzed by chi-square tests and could also employ blocking variables.

In addition, the interrelationships of the various criterion measures could be examined both within and across institutions.[5] It is possible to compute correlations between any pair of variables recorded for each student (e.g., personal characteristics, achievement scores, and satisfaction indices). It may also be of interest to examine more complex relational structures (e.g., the most highly interrelated patterns of achievement and satisfaction, differences among these patterns both for the various institutions and for different ability levels of the students).

(8) Prepare report. Summarize findings, discuss implications and limitations of data, suggest problems and areas for further research, outline approaches which appear to be most fruitful for future assessment studies. — November through February

An example of the kinds of data which could be presented in the project report is given below. The Undergraduate Program (UP) Natural Science Area test (formerly the GRE Institutional) is used as an illustrative criterion measure. The table would provide information bearing upon question 4 above.

TABLE B

Mean UP Natural Science Scaled Scores for State University and State College Seniors at Four Levels of Academic Ability

Ability Level	State University System			State College System			
SAT National Percentile Rank	Campus A	Campus B	Campuses A and B (Combined)	Campus C	Campus D	Campus E	Campuses C,D,E (Combined)
90-100	715	685	700	660	670	665	665
75- 89	660	640	650	640	655	640	645
50- 74	585	575	580	580	590	585	585
below 50	450	450	450	465	475	470	470

It is evident from this hypothetical table that the university campuses are more effective for the highest ability (top 10%) students, while the state colleges are more effective for lower ability (bottom 50%) students. Between these two ability categories, there are no important inter-system differences. Why this should be so (if indeed it were a real finding) would require a synthesis of several types of data.

Some intrasystem differences appearing in this table are also noteworthy. For example, seniors above the 75th percentile in academic ability at University Campus A are quite superior to seniors of the same ability level at Campus B in their performance on this criterion measure. At lower ability levels, these differences disappear. Within the state college system, Campus D is consistently somewhat more effective than either of the other two across the entire range of student ability.

Plan B. Assessment of Sophomores: General and/or Vocational Education

The assessment plan for the first two years of college consists of two parts, focused separately at two somewhat disparate student populations--those in general (including two-year college transfer) education programs and those in terminal vocational programs. Clearly, the two groups are not appropriately evaluated on the same criteria.

Evaluation of the general (transfer) programs at the two-year colleges would be a variant of Plan A (for four-year institutions) just presented. Plan B permits comparison, on academic and intellectual dimensions, of graduating general education students across two-year colleges, of sophomores across four-year institutions, and between systems of two-year and four-year colleges.

Assessment of vocational education programs by direct measurement of student learning is not recommended because suitable tests, which would ideally be criterion-referenced or performance-based, are generally not yet available.[6] Assessment for vocational areas would be thus largely through self-report and directed at perceived effectiveness or quality of the training program, as well as satisfaction with other aspects of the community college experience.

Except for those concerning acquisition of specialized (major field) knowledge, many of the same kinds of questions raised under Plan A--for seniors--may also be answered with respect to sophomores by the Plan B assessment. Questions 1, 3, 4, 5, and 7 under Plan A would be applicable to Plan B. Or, put somewhat differently:

(1) Are there differences from one campus to another in the "general education effectiveness" of the first two years? In intellectual disposition and/or satisfaction with college, after the first two years?

(2) Are there differences on the intellectual competence dimensions between end-of-year sophomores at two-year and four-year institutions? That is, for example, in states with both two-year and four-year systems, is one

system more "general education effective" than the other, with student ability taken into account?

(3) Are there differences in "vocational education effectiveness" (all such programs combined) from one community college to another?

(4) For specific programs, e.g., cosmetology, auto mechanics, and so forth, are there differences in (student-perceived) effectiveness from one college to another?

Because of the general similarity to Plan A in approach, criteria and measuring instruments, and data processing and analysis, the outline for conducting the Plan B assessment which follows points out only its unique or special aspects.

(1) Determine the assessment criterion variables and measuring instruments. October through March

General Education Students: With the exception of specialized knowledge, the same criterion variables and measures as those listed under Plan A are suggested. If Plans A and B were undertaken simultaneously, this would allow examination of trends and provide comparative data for sophomore and senior level assessment.[7]

Vocational Education Students: A specially prepared Vocational Student Questionnaire would cover such items as perceived quality of teaching, equipment, program/course organization, interaction between vocational and general education student groups, employment advising, job prospects, etc.

(2) Prepare Sophomore Questionnaire. This would cover the same kind of content as the Senior Questionnaire (Plan A), appropriately adjusted for the difference in levels. Three somewhat different forms might be required (for the sophomores at four-year institutions, the general/transfer students at two-year colleges, and the vocational education students), depending upon the structure of the state's higher education complex. The content would, of course, be overlapping; most items would appear on all three forms, some on only two forms, and a few on only one form. October through March

(3) Determine information to be used for control of differential input. Because most two-year institutions have open admissions, there are no commonly used entrance tests. Possible substitutes include equated scores on several widely used tests (SCAT, CQT, CGP, for example), or high school grades or class rank. October through March

(4) Design data management procedures. Same as Plan A. October-March

(5) Conduct the assessment. Same as Plan A. April

(6) Process data. Same as Plan A. May-June

(7) Analyze data. As with Plan A, the sorts of questions of interest to the assessment and system staff will determine the specific analyses to be carried out. Many of the suggested comparisons and breakdowns for seniors can be applied analogously to sophomores.[8] July through September

(8) Prepare report. Same as Plan A. October-February

Plan C. Cross-Sectional Assessment: Focus on Intellectual and Nonintellectual (Affective) Development

Numerous observers of higher education have recognized that academic accomplishment is not the sole and perhaps not even the most important objective of a college education. They emphasize that the college experience should enhance growth and development in the non-cognitive domain as well--that a person should emerge from college psychologically integrated, interpersonally competent, socially responsible, and generally effective in the conduct of his everyday affairs.

Plan C provides a means of assessing differential impact of various institutions upon such nonintellectual areas, as well as upon academic achievement. The assessment of effectiveness on intellectual dimensions would proceed along the same lines as that described in Plans A and B, providing, in addition, a means of examining ability level differences both for incoming freshman classes and for dropouts over the three-year time period. Plan C also permits determination of the relational structure of intellectual and nonintellectual variables for freshmen, which is free of any institution effect and thus provides a basis for evaluating (possibly) different relational patterns at the end of the sophomore and senior years.

The assessment for the nonintellectual variables would be carried out by comparisons of cross-sectionally derived patterns of change. We presume that there will be no pre-admission scores on these sorts of measures available for the current sophomores and seniors, so that a longitudinal growth study would not be possible within the projected time schedule for the assessment.[9] It is not anticipated, however, that the distribution of incoming students according to nonintellectual attributes will be substantially different from year to year (over the short time period of at most three years). Thus it will be possible both to determine relationships between the criterion variables and dropping out, and to analyze separately the dropout and continuing student data at the end of the freshman and sophomore years (the highest dropout probability period). These conditions provide a reasonable basis for the appropriateness of cross-sectional comparisons.

Plan C, the most comprehensive of the four, enables answering all the kinds of questions posed under Plans A and B. The one exception, as Plan C is drawn, pertains to assessment of vocational education, and this could be accomplished by adding the (Plan B) vocational component to the Plan C sophomore assessment.

Additionally, Plan C allows examination of a variety of questions concerning the <u>progress</u> of intellectual and affective change during the undergraduate years. For example:

(1) To what extent does <u>general</u> academic learning occur during the first two years, rather than the last two? At particular colleges? From one system to another?

(2) What is the pattern of differential preparedness for upper division work in various major fields, as indexed by end-of-sophomore-year performance on subject field examinations?

(3) What is the pattern of development of intellectual attitudes and commitments? Does such commitment, for example, tend to occur earlier at some colleges than at others?

(4) Do seniors tend to be more, or less, satisfied with their college work than end-of-year sophomores? Are there differences from one institution to another?

(5) In the nonintellectual (affective) domain, are there differences from one campus to another on measures of attributes such as Personal Integration, Social Conscience, and Self Regard? Between graduates of public and private (four-year) systems? Between end-of-year sophomores in four- and two-year systems?

(6) What is the pattern of change in these attributes during the undergraduate years--from the time of freshman entry, to the end of the sophomore year, to the time of graduation? At specific colleges? Throughout the system?

(7) What are the relationships between designated student background factors--academic ability, socioeconomic level, sex, for example--and the various nonintellectual criteria?

(8) What institutional/program characteristics are associated with high (or low) scores on the affective measures? Do graduates in the humanities, for example, score relatively high on the measure of "Self-Actualizing Value"? Do sophomores who have lived on campus score higher than commuters on the (hypothetical) measure of interpersonal competence?

A general outline for the Plan C assessment procedure is as follows.

(1) Determine the criterion variables and specific instruments for assessing each. An illustrative set of variables is presented in Table A. Since the instruments and variables for all three classes would not overlap entirely, it will be necessary to coordinate choices of certain ones (e.g., goals and expectations of entering freshmen and satisfaction indices of sophomores and seniors), and to consider the kinds of data which may validly be reported retrospectively by sophomores and seniors (e.g., original goals, changes, etc.). April-August

(2) Develop the Questionnaires. The information requested would cover such things as background factors, goal expectations and attainment, perceptions of college programs, and other data bearing on questions of interest to the system. Somewhat different forms will be needed for each class, each requiring approximately 1/2 hour to complete.

Freshman Questionnaire — July-August
Sophomore and Senior Questionnaires — October-March

(3) Determine information to be used for control of differential input. Same as Plan A. October-March

(4) Design data management procedures. Same as Plan A. August-March

(5) Conduct the assessment. The same general considerations discussed under Plan A with respect to sampling and testing situation apply here. Because of the longer testing time required for Plan C, consideration may be given to decreasing it somewhat by distributing the tests over different samples of each class, but the extent to which this can be done is limited by the class size and the number and nature of breakdowns to be analyzed.

Freshman assessment — September
Sophomore and Senior assessment — April

(6) Process data. Same as Plan A. — May-June

(7) Analyze data. In general, the analyses would follow those described for Plan A, with inclusion of cross-sectional comparisons. — July through October

An additional type of effectiveness assessment could be provided by analyses of dropout data. Dropouts would be identified by comparing registration lists for the next two semesters (three quarters) with that of the entering freshmen, and for the next semester (quarter) with that of the sophomores. It will be possible to ask sophomores at the time of testing (on the Questionnaire) if they intend to return in the fall and if not, why. This information could also be obtained from the freshman dropouts by a mailed questionnaire. Frequencies of students dropping out for various reasons can then be compared across institutions by chi-square analyses.

(8) Prepare report. Same as Plan A. — November-May

Plan D. Survey of Recent Alumni

A number of readily meaningful indices of institutional effectiveness insofar as students are concerned would derive from a systematic survey of recent alumni--their employment status and satisfaction, various civic activities, diverse cultural interests and activities, perceptions of various college experiences, and so forth. As with the other assessment plans, the survey must be differentially comparative, in the sense of gathering the same information from alumni of different institutions, and it should provide a means for taking into account differential academic ability.

A system-wide alumni survey could shed light on a host of policy-relevant

questions. Some of these include:

(1) To what extent are recent graduates from the various colleges--in total and by subject field--finding employment?

(2) To what extent do they regard their employment as personally satisfying? Consistent with college studies? Consistent with perceived level of intellectual ability and/or training?

(3) What changes are seen by alumni as necessary to bring about a better fit between system (and college) curriculum policies and practices, and current job-market realities? What changes may be needed in view of estimated shifts in job markets?[10]

(4) To what extent are graduates entering postgraduate programs--from the total system, by campus, and by major field and sex?

(5) Which graduate programs--in the same system, other systems in the state, private universities in and out of state--are receiving the system's graduates?

(6) What are the reasons graduate students are in particular programs, and what suggestions do they have for the system in question for modifying its graduate programs to better meet the graduate school needs of its alumni?

(7) Some of the same questions may be asked concerning transfer students from two-year colleges--where they go and why, articulation difficulties, suggested improvements of transfer programs, and so forth.

(8) Broad educational policy questions: What is the appropriate mix of liberal/general education and specialized occupational training? What modes of instruction are perceived to be most effective? Should mastery of designated content and/or skills be required for the degree (or certain degrees)?

(9) What are alumni doing with their lives outside the occupational and educational spheres? What are their interests and activities in, let us say, the cultural, political, community service, and recreational domains?

(10) What are some of their attitudes and opinions: About the general quality of their lives? About their future prospects? About particular social and political institutions? About specific problems and issues--environmental protection, population planning, the role of science, corruption in government, for example?

Our suggested general procedure is outlined in the following steps:

(1) Prepare the survey questionnaire.[11] A variety of content could be considered for inclusion: present circumstances (graduate school, employment, etc.), job satisfaction, earnings, community (e.g., service) activities, and so forth, as indicated in Table A. — September through March

The questionnaire should be brief--perhaps a cover and three pages of questions printed on a single 8 1/2 by 17 inch sheet (folded). Print questionnaire to be compatible with optical scanning equipment. Card-punching response data would be economical only for small systems--up to four or five institutions or three or four thousand respondents.

(2) Determine survey population. Identify (produce a list of) all bachelor degree recipients (or AA and certificate recipients, at two-year colleges) two years prior to the time the survey is to be conducted, e.g., graduating seniors in May, 1972. — October

(3) Determine the survey sample. From the above population, form a stratified random sample of 2000 alumni. At institutions where the graduates numbered fewer than 2000, survey the entire class.[12] Stratify the sample by sex and by general academic major (education, social sciences, business, etc.). — November

(4) Locate individuals in sample. Determine the present address of the 2000 individuals. Use all sources available--alumni office, placement service, department personnel, possible friends, etc. For untraceable individuals, select replacements at random from the appropriate sex-major field cells. — December through February

(5) Determine academic ability score for sample subjects. Same as Plan A. — October-March

(6) Computerize names/addresses. Develope magnetic tape or addressograph plates with names and addresses for efficient addressing of survey envelopes and follow-up postcards. — March

(7) Mail survey package. (Envelope, questionnaire, return envelope.) — April

(8) Follow-up. Mail "broadcast" postcard one week later to entire sample, urging cooperation and advising individuals who have responded to disregard the card. — April

(9) Process returns. Edit and code (any open-ended questions) as returns come in (coding systems must be standard for returns for all campuses). — May-June

(10) Data Processing. Transcribe responses from questionnaires via optical scan equipment or key punch. — July

(11) Data Analyses. Tabulate responses by frequency and percent for: — August

(a) the total sample (all returns) from each college;[13]
(b) major field by sex (and possibly other) breakdowns, for each college;
(c) all respondents from all campuses in the system, aggregated in total and by major field/sex--to understand deployment of graduates from the total

system, possibly in comparison with other state systems, and with available national data (census, Gallup Poll, etc.).

(12) Prepare project report. Summarize findings, give possible reasons for differential patterns by college, set forth implications. September through January

* * * *

We have attempted in this paper to respond in a tangible and practical way to the increasing call for evidence of institutional effectiveness. Four plans are proposed, representing different, but related and complementary, approaches to the question of effectiveness assessment. The problem is addressed from different time perspectives in the educational process (i.e., sophomores, seniors, alumni), different types of criteria employed (academic, vocational preparation, affective, etc.), and different levels of post-secondary education institution/program (two- and four-year institutions, academic and vocational programs). Each plan is based on an interaction paradigm intended to determine areas of differential institutional effectiveness in the context of student-institution "fit".

We have offered these ideas for institutional assessment in the belief that they can generate potentially useful information about the effectiveness of campuses and groups of campuses, that people on the campuses and in the systems will consider the required investment of effort and resources worthwhile, that cooperative multi-constituency planning and execution of the assessments (and the consequent enhanced legitimacy of the findings) is possible, and that the entire undertaking can indeed lead to institutional renewal and to full realization of the educational goals of every student.

Notes

[1] A state university, four-year college, or community college system; or the private colleges in a state, region, or consortia.

[2] While some conception of student learning/development would be a fundamental goal at almost all institutions, many campuses would attach importance to other goals as well. Thus a comprehensive university would wish to construe effectiveness in terms of research and scholarly contribution, and perhaps public service. Indeed, numerous additional effectiveness criteria are conceivable; Peterson (1971), for example, suggested responsiveness (to community educational needs) and general campus morale as (additional) indices of institutional effectiveness.

[3] Since students cannot be randomly assigned to different colleges, statistical (i.e., randomization-based) corrections for differential student input characteristics are precluded. There simply is no way to generate an "expected" score which is not institution-bound. Furthermore, if there are student-institution interaction patterns, there is no sense in which such a removal of input characteristics has meaning in the assessment of institutional effect. Thus, even if it were clear how to go about it, to do so leaves one with a nagging sense of unreality. One may well wonder what it means to say, for example: If all students were of the same (average) ability level, institution A would be most effective in developing the academic potential of its students.

We suggest that the most appropriate manner of handling the differential spread of student talent (of all sorts), aspirations, cultural backgrounds, and so forth, is to capitalize on their presence to identify optimal student-college matchups. In this way the diversity of institutional goals and programs as well as their unique strengths may be recognized.

[4] The one-by-one assessment of criterion variables may be augmented by the multivariate extension of ANOVA (MANOVA) to examine patterns of effectiveness across several criteria. A routinely generated by-product of MANOVA is a discriminant function analysis (set of regression functions which maximally, and orthogonally, differentiate the student groups) which would identify particular patterns and levels of differential effectiveness.

It may also be of interest to create student typologies (based on combinations of ability, degree aspiration, background factors, etc.) by classification on latent dimensions derived by factor analysis. Factor scores may be estimated and used as a taxonomic basis for comparisons on any or all of the effectiveness criteria, singly and/or in combination.

[5] Is satisfaction, for example, related to academic achievement? To intellectual disposition? What is the relationship between breadth and depth in academic performance? Is it the same at all institutions? At all ability levels?

[6] A broad range of vocational competency tests will soon be made available through the Center for Occupational and Professional Evaluation and the National Occupational Competency Testing Institute, both administered by ETS.

Notes (continued)

[7] For example, it would permit evaluation of the transfers vs. natives findings at the senior level with comparisons of prospective transfers with natives at the time of transfer, zeroing in on just when the differences (if any) come about.

[8] Some sorts of interrelationships of criterion variables would be of particular interest for the two-year colleges, and especially how those relationships are affected by whether one is a general or vocational education student. For example, is the relationship between campus climate and goal fulfillment greater for one group than for the other? At some campuses more than others?

[9] It should be understood, however, that a choice between longitudinal and cross-sectional designs is not entirely a tradeoff of legitimacy of method vs. expediency of execution. There are clearly some advantages to having input data (it is difficult to make a case for not wanting more data), but, despite oft-stated claims to the contrary, the longitudinal method is not so clean as its proponents profess. The major criticism of cross-sectional designs is that one cannot be certain that the current freshman class is representative of the present sophomore, junior, and senior classes at the time they were freshmen, with respect to the criterion being measured. This would primarily be due to: (1) dramatically different freshman student bodies in the different years (which is unlikely to be very much of an issue in the proposed design since no school explicitly selects on such criteria and the classes tested are only one and two years apart), or (2) the presence of prospective dropouts in the freshman class which will bias the comparisons with other classes if dropping out is substantially related to the criterion variable (a situation which can be partially controlled in the proposed design; see text).

However, exactly these same problems also restrict the generalizability of the results of longitudinal studies, which must assume that the magnitudes and kinds of changes which occur for one student body in one time period can be validly extended to other student bodies at other points in time. But a single graduating class may not be at all representative of incoming classes several years later (after, for example, the results of a four-year study are analyzed, reported, digested, and acted upon). The characteristics of incoming student bodies and/or dropouts may be markedly affected by the gradually changing nature and policies of the institutions themselves (e.g., open admissions), and by the differences in goals and interests of students, both those responsive to personal decisions and those induced by societal changes (e.g., shifting job markets). These uncertainties suggest that, from a policy-making point of view at least, the timeliness of cross-sectional results outweighs the extra control on variability which a longitudinal study permits.

[10] Repeated alumni surveys, say at three-year intervals, would permit noting trends that could be useful in college and system long-range planning.

Notes (continued)

[11] An excellent prototype is contained in Perrella (1973). Use of questions from recent national surveys, needless to say, allows comparing local findings with national data.

[12] The large number is required in view of the expected 40 to 60 percent return rate.

[13] Modern survey research commonly employs weights to correct for sampling bias, a frequently cumbersome procedure which is not always cost-effective. In the survey outlined here, weights could be applied to correct for differential return rates by sex and major field.

References

1. Carnegie Commission on Higher Education. Continuity and Discontinuity: Higher Education and the Schools. New York: McGraw-Hill, 1973.

2. Perrella, Vera C. Employment of Recent College Graduates. Special Labor Force Report 151. Washington, D. C.: U.S. Department of Labor, Bureau of Labor Statistics, 1973.

3. Peterson, Richard E. College Goals and the Challenge of Effectiveness. Princeton, N.J.: Educational Testing Service, 1971.

4. Peterson, Richard E. Intellectual Competence: Definition and Measurement. Research Memorandum 71-15. Princeton, N.J.: Educational Testing Service, 1971.

SUMMARY

Strategies for Assessing Differential Institutional Effectiveness in Multi-Campus Systems (Peterson and Vale, ETS Berkeley, October 1973)

This paper is an attempt to respond to the need for workable procedures for assessing the effectiveness of programs and institutions in complex postsecondary education systems. Effectiveness is taken to mean the capacity of the institution to advance student development--academic, vocational, and affective. Four general strategies are outlined in the paper; each is intended to:

(1) Yield information directly applicable to policy issues and decisions;

(2) Yield information in a timely fashion;

(3) Be implementable, in the sense of practical feasibility.

In brief summary, the four plans are the following:

Plan A. Senior Assessment: Intellectual Competence. This plan focuses on the development, after four years of college, of a number of intellectual and academic attributes, while taking account of the general academic ability of the student at the time he enters as a freshman.

Plan B. Sophomore Assessment: Intellectual and/or Vocational. Plan B affords a method for evaluating institutional effectiveness in either general (and transfer) education or vocational training programs (or both) during the first two postsecondary years (again, accounting for differential ability level).

Plan C. Cross-sectional: Intellectual & Nonintellectual. Plan C generates all the information yielded by Plans A and B (which focus on intellectual criteria), plus a cross-sectional description of entering freshmen, end-of-year sophomores, and graduating seniors on nonintellectual (affective) as well as intellectual criteria.

Plan D. Alumni Survey. The criteria of effectiveness in this plan are various achievements and activities on the job, in graduate school, or elsewhere for alumni two years after graduation (from either a two- or four-year institution). As with the other plans, this one also permits taking freshman ability into account in assessing post-graduate achievements.

Each plan is discussed separately, from the standpoints of (1) general purpose and logic, (2) illustrative policy questions answerable from the assessment, and (3) steps involved in conducting the assessment, together with a suggested time schedule.

While the four strategies assume cores of common goals--and therefore assessment criteria--across institutions, all the ideas set forth in the paper are regarded as flexible, as adaptable to system resources, policy interests, and information needs. Cooperative multi-institution/multi-constituency planning for and execution of the assessments, including, in particular, definition of criterion variables and choice of instruments, are taken for granted.

A general data analysis approach is proposed which emphasizes appraisal of: (1) student growth in the sense of value added; and (2) interactions between student and program characteristics, with potential for identifying effective student-institution "fits." This approach, while requiring assessment of relatively large numbers of students, avoids many of the statistical and interpretational pitfalls frequently encountered in evaluation studies in education, as well as providing a systematic basis for program renewal to better accommodate the varied interests of a diversified student population.